SOLVATION FREE ENERGY PREDICTIONS FROM MOLECULAR DYNAMICS SIMULATIONS BY IMPROVED ALCHEMICAL PATHWAYS AND OPTIMIZED FORCE FIELD PARAMETERS

A. Mecklenfeld[1,2], G. Raabe[1*]

[1] Institut für Thermodynamik, TU Braunschweig, Hans-Sommer-Str. 5, 38106 Braunschweig
[2] Center of Pharmaceutical Engineering, TU Braunschweig, Franz-Liszt-Str. 35a, 38106 Braunschweig

g.raabe@tu-braunschweig.de

ABSTRACT

The rapid development of affordable medications relies on the knowledge of suitable solvents for potential active pharmaceutical ingredients. Molecular simulations can be interpreted as computational experiments and may complement laboratory experiments, as they not only enable the calculation of thermophysical properties but also allow for an insight into a systems behavior on the molecular level. Relative solubilities can be predicted by using molecular simulations to calculate the Gibbs free energy of solvation ΔG_{solv}. We developed algorithms to reduce the required computational effort and increase the statistical precision by improved alchemical pathways. However, the agreement between ΔG_{solv} predictions from simulations and experimental data depends on the molecular models (force fields) used. To consider polarization effects, we combined partial charges derived by the IPolQ-Mod method with the General Amber Force Field (GAFF) and found a comparable ΔG_{solv} accuracy to GAFF and its default RESP charges for a large set of compounds in various solvents. We initiated a parameter optimization to improve the accuracy of GAFF/IPolQ-Mod and present our current results of the refitting process.

Keywords: Molecular simulation, solvation free energy, force field optimization

INTRODUCTION

Molecular simulations, e.g. with the Gromacs (Abraham et al. 2015) software, can complement laboratory experiments, as they not only allow for the calculation of thermophysical properties but also for the qualitative assessment of a systems behavior on the molecular level. A property of special interest is the Gibbs free energy of solvation ΔG_{solv}, which is fundamental for the calculation of e.g. relative solubilities, partition coefficients or activity coefficients. ΔG_{solv} characterizes the isothermal isobaric change of states of a solute molecule that transitions from a vacuum into a condensed phase. The corresponding change in free energy describes the strength of the solute/solvent interactions as well as the impact of entropic effects. For accurate free energy results, the thermodynamic end states, i.e. solute in vacuum and solvent phase respectively, need to exhibit sufficient configurational space overlap. In practice, a linking chain of intermediate states with scaled solute/solvent interactions, the so called alchemical pathway, is constructed. It allows for overlap between neighboring states and partial free energy results are summed for the total outcome. However, while a large number of intermediates enables good overlap and high statistical precision, the required computational effort increases with the number of intermediates.

https://doi.org/10.24355/dbbs.084-202001221308-0

On the other hand, ΔG_{solv} results strongly depend on the molecular models (so called force fields) applied. Widely used force fields (Wang et al. 2004; Vanommeslaeghe et al. 2010; Jorgensen et al. 1996) for the description of drug-like molecules assign fixed partial charges to the atomic sites. Consequently, these force fields are not capable to consider important polarization effects. For an at least implicit representation of polarization effects, we combined the IPolQ-Mod (Muddana et al. 2014) partial charge calculation method with the General Amber Force Field (GAFF) (Wang et al. 2004) and compared the ΔG_{solv} results to those obtained with GAFF but with the default RESP (Bayly et al. 1993) charge calculation method. In a previous study (Mecklenfeld und Raabe 2017a), we found comparable agreement between results with GAFF/RESP and GAFF/IPolQ-Mod to experimental data and concluded that GAFF/IPolQ-Mod is a suitable basis for the application of the physically motivated IPolQ-Mod method. However, as the alteration of the charge calculation method perturbs the self-consistency of the GAFF model, we also proposed the optimization of relevant atom types in various chemical environments in order to improve the ΔG_{solv} predictions.

RESEARCH CONCEPT

OPTIMIZATION OF THE NUMBER AND DISTRIBUTION OF INTERMEDIATES STATES

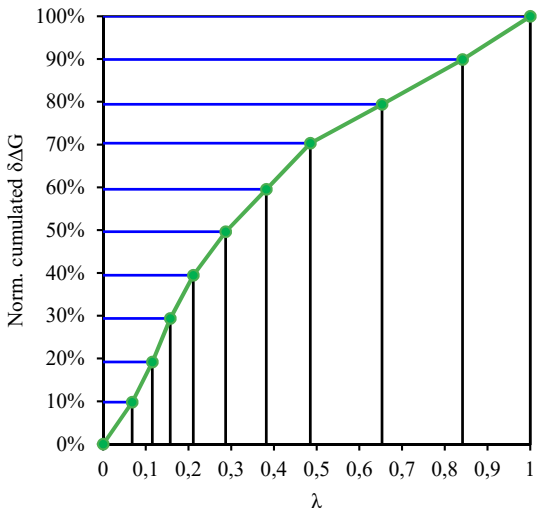

Fig. 1. Curve progression of normalized cumulated partial uncertainties with regard to λ. At the end of the λ state optimization, the objective of equal partial uncertainties, represented by equidistantly spaced cuts with the ordinate axis (blue lines), is well meet.

Intermediates states are characterized by λ-scaled solute/solvent interactions, whereas λ=0 equals no solute/solvent interactions (solute in vacuum) and λ=1 corresponds to full interactions in the condensed phase. By defining a vector of λ values between 0 and 1, configurational space overlap can be established. The solvation free energy difference is thereby split into partial results $\Delta G_{solv,i,i+1}$ between states λ_i and λ_{i+1}, which are summed for the total outcome. A priori, a suitable number or distribution of λ states is unknown and thus sufficient configurational space overlap is often accomplished by using a large number of intermediates.

As each partial result $\Delta G_{solv,i,i+1}$ contributes to ΔG_{solv}, our optimization goal is to ensure even partial uncertainties $\delta\Delta G_{solv,i,i+1}$ for improved statistical precision of the overall ΔG_{solv} result. We developed algorithms that allow for the adaptation of the number and distribution of the intermediates by a series of short trial simulations (Mecklenfeld und Raabe 2017b). As illustrated in Figure 1, cumulated partial uncertainties are interpreted as a function of λ. Using interpolation, new λ states can be defined to ensure the same level of statistical precision for all intermediate results. The configurational space overlap is analyzed after each of the short simulations and empirically derived algorithms adapt the number of intermediates if the overlap is considered to be insufficient or too excessive. After the number and distribution of intermediates converged, the actual production phase is initiated.

CALCULATION OF PARTIAL CHARGES

The IPolQ method is an abbreviation for “Implictly Polarized Charges” and was proposed by Cerutti et al. (Cerutti et al. 2013). The approach was later simplified by Muddana et al. and named IPolQ-Mod accordingly. Using ab initio calculations in Gaussian (Frisch et al. 2009), molecule geometries are optimized at the HF/6-31G* level of theory. Partial charges for the solute molecule are calculated both in the vacuum and condensed phase at the MP2/aug-cc-pVDZ level of theory and charges from both phases are then averaged. For the description of charges in the solvent, the PCM model by Mennucci et al. (Mennucci et al. 1998) is employed. For default GAFF charges, the HF/6-31G* level is applied. All partial charges are derived by a two-level RESP scheme.

https://doi.org/10.24355/dbbs.084-202001221308-0

OPTIMIZATION OF ATOM TYPES

For the optimization of Lennard-Jones parameters ε_{ii} and σ_{ii} for relevant atom types, we employed the robust Nelder-Mead (Nelder und Mead 1965) algorithm and use an objective function that considers weighted normalized root mean square deviations for solvation free energies and densities of the pure compounds. To reduce the computational effort, we utilize a thermodynamic cycle for the parameter optimization process. The training data set includes 358 solvation free energy results and 286 densities in a broad temperature range for a multitude of model compounds and chemical environments to account for prominent functional groups in potential drug candidates.

RESULTS AND DISCUSSION

In Figure 2, solvation free energies from GAFF/RESP and GAFF/IPolQ-Mod are plotted versus the corresponding experimental data. The figure demonstrates comparable accuracies for the different charge calculation methods, though specific outliers exists for both.

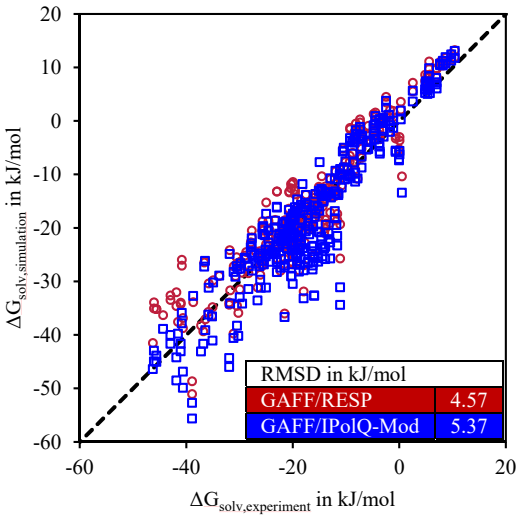

Fig. 2. Solvation free energy results from GAFF/RESP and GAFF/IPolQ-Mod over experimental data.

Simulation results for the ongoing optimization of atom types (ch2, ch3, ca, oh, ohP, os and cl already considered) are depicted in Figure 3. The figure highlights the improved agreement between ΔG_{solv} results using GAFF/IPolQ-Mod + LJ-refit and experimental data compared to the GAFF/RESP and GAFF/IPolQ-Mod models.

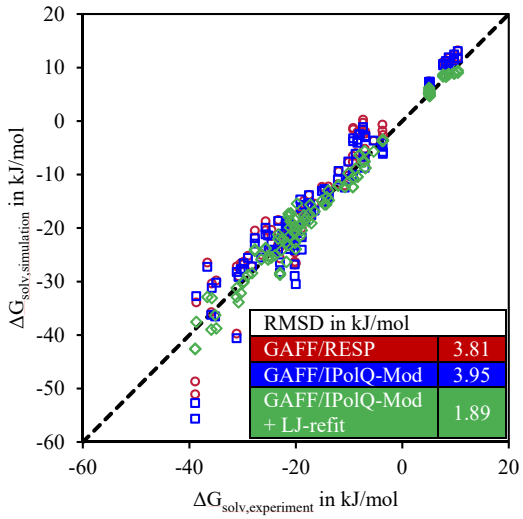

Fig. 3. Solvation free energy results from GAFF/RESP, GAFF/IPolQ-Mod and GAFF/IPolQ-Mod with refitted Lennard-Jones parameters over experimental data.

CONCLUSIONS

Molecular simulations are capable to complement laboratory experiments. The precise prediction of solvation free energies for the estimation of relative solubilities requires sufficient configurational space overlap for the ΔG_{solv} results to converge, and accurate molecular models for the description of the molecule interactions. We have developed methods to ensure statistical precise free energy results with reduced computational effort, and studied the impact of IPolQ-Mod partial charges for an implicit representation of polarization effects on the GAFF model compared to the default force field charges.

Though IPolQ-Mod adds physical details to the model, the charge scheme perturbs the self-consistency of the force field. We therefore optimize Lennard-Jones parameters of GAFF atom types for IPolQ-Mod partial charges, and our results demonstrate significantly improved agreement with experimental data.

Suitable molecular models for exact predictions of solvation free energies and linked thermophysical properties may improve the understanding of solubility processes in order to reduce the effort for the development of medications.

https://doi.org/10.24355/dbbs.084-202001221308-0

ACKNOWLEDGMENT

The study is funded by a Georg-Christoph-Lichtenberg grant. All simulations were performed on a cluster of the North-German Supercomputing Alliance (HLRN). We greatly appreciate the support.

REFERENCES

Abraham, Mark James; Murtola, Teemu; Schulz, Roland; Páll, Szilárd; Smith, Jeremy C.; Hess, Berk; Lindahl, Erik (2015): GROMACS. High performance molecular simulations through multi-level parallelism from laptops to supercomputers. In: *SoftwareX* 1-2, S. 19–25. DOI: 10.1016/j.softx.2015.06.001.

Bayly, Christopher I.; Cieplak, Piotr; Cornell, Wendy; Kollman, Peter A. (1993): A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges. The RESP model. In: *J. Phys. Chem.* 97 (40), S. 10269–10280. DOI: 10.1021/j100142a004.

Cerutti, David S.; Rice, Julia E.; Swope, William C.; Case, David A. (2013): Derivation of fixed partial charges for amino acids accommodating a specific water model and implicit polarization. In: *J. Phys. Chem. B* 117 (8), S. 2328–2338. DOI: 10.1021/jp311851r.

Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R. et al. (2009): Gaussian 09, Revision A.02. Version. Wallingford CT: Gaussian, Inc.

Jorgensen, William L.; Maxwell, David S.; Tirado-Rives, Julian (1996): Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. In: *J. Am. Chem. Soc.* 118 (45), S. 11225–11236. DOI: 10.1021/ja9621760.

Mecklenfeld, Andreas; Raabe, Gabriele (2017a): Comparison of RESP and IPolQ-Mod Partial Charges for Solvation Free Energy Calculations of Various Solute/Solvent Pairs. In: *J. Chem. Theory Comput.* 13 (12), S. 6266–6274. DOI: 10.1021/acs.jctc.7b00692.

Mecklenfeld, Andreas; Raabe, Gabriele (2017b): Efficient solvation free energy simulations. Impact of soft-core potential and a new adaptive λ-spacing method. In: *Mol. Phys.* 115 (9-12), S. 1322–1334. DOI: 10.1080/00268976.2017.1292008.

Mennucci, Benedetta; Cammi, Roberto; Tomasi, Jacopo (1998): Excited states and solvatochromic shifts within a nonequilibrium solvation approach. A new formulation of the integral equation formalism method at the self-consistent field, configuration interaction, and multiconfiguration self-consistent field level. In: *J. Chem. Phys.* 109 (7), S. 2798–2807. DOI: 10.1063/1.476878.

Muddana, Hari S.; Sapra, Neil V.; Fenley, Andrew T.; Gilson, Michael K. (2014): The SAMPL4 hydration challenge: evaluation of partial charge sets with explicit-water molecular dynamics simulations. In: *J Comput Aided Mol Des* 28 (3), S. 277–287. DOI: 10.1007/s10822-014-9714-6.

Nelder, J. A.; Mead, R. (1965): A Simplex Method for Function Minimization. In: *Comput. J.* 7 (4), S. 308–313. DOI: 10.1093/comjnl/7.4.308.

Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J. et al. (2010): CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. In: *J. Comput. Chem.* 31 (4), S. 671–690. DOI: 10.1002/jcc.21367.

Wang, Junmei; Wolf, Romain M.; Caldwell, James W.; Kollman, Peter A.; Case, David A. (2004): Development and testing of a general amber force field. In: *J. Comput. Chem.* 25 (9), S. 1157–1174. DOI: 10.1002/jcc.20035.

https://doi.org/10.24355/dbbs.084-202001221308-0